Modelling the Dynamics of Musical Engagement

Jonathan Impett

Faculty of Music
University of Cambridge
West Road
Cambridge CB3 9DP
UK
jfi21@cam.ac.uk

Abstract

This paper presents work in progress on the modelling of emergence in an intelligent music composition/performance system. A model is described which seeks to encourage such emergence by facilitating interaction between musician and system. A complex adaptive systems population of musical behaviours inhabits a landscape which in turn is shaped by a process of description and redescription of the events which constitute its environment. It is suggested that the mutually evolving space thus generated might be considered a useful model of the emotional engagement between musician or listener and an external musical "object".

Introduction

The system described in this paper forms part of an ongoing project looking at ways of modelling engagement and emergence in interactive music systems; computer-based systems which compose and perform in real time together with human musicians. The composition consists of the environment of rules, processes and mechanisms by which the computer calculates its response to data derived from the musicians' performance. The current system is centred around the "meta-trumpet" described in (Impett 1994), which generates various forms of physical sensing information (position, movement, pressure etc.), as well as that representing the sound of the instrument (note name, pitch bend, volume), all as MIDI data. The compositions are implemented in C++ on a Silicon Graphics O2, which also performs the sound synthesis and processing output.

An understanding of the cognitive and emotional aspects of music as rooted in a process of description and prediction goes back at least to (Meyer 1956). More recent theories offer a more precise model of abstracted (Narmour 1990) or style-specific (Lerdahl and Jackendorf 1983) cases, but generally do not tackle the dynamically hierarchical nature of this experience pointed to by Meyer. That cognised structure and emotional response can vary substantially even between performances of the same material points to the fact, as Meyer insists, that music is a highly time-dependent and contextual, situated activity. The multi-level and non-referential nature of music make it ideally suited to the study of such phenomena; indeed it could be argued that the importance of music lies precisely in its affording the possibility of self-study of the structure and dynamics of emotion.

A Framework - Musical Activity, Engagement and Emergence

Whilst the present project deals specifically with musician/computer interaction, engaged musical activity - attending to music - is here understood in a wider sense, incorporating composition, performance, listening and analysis. The focus of this study is therefore not a model or simulation of some given musical surface, but an object of musical engagement; the "other" in musical activity in general, and the musical object in the terms of this paper. The musical object will be seen to have a dual nature. It both constitutes an extension of the musical faculty of the person with whom it interacts, and has a degree

Any technology that would engage with human experience implies a cognitive theory. Here it is suggested that musical listening parses its input in a multi-levelled, hierarchical fashion by identifying and predicting strands of behaviour. It shares this mechanism with "listening-in-the-world", but constrains both the acoustic and intentional spaces. This could be considered in terms of Bregman's search for regularities in the acoustic scene (Bregman 1993), constraining both the type of behaviours sought and the significance ascribed to them according to our experience of the present type of musical object. The behaviours of most interest will be neither fully predictable nor completely random, but exhibit dynamic and occasionally emergent behaviour, as is the nature of natural systems. Mutual change in behaviour patterns - interactive emergence - is therefore the object of interest, generally the result of the interaction between behaviours, with their environment, or with themselves. Such a cognitive mechanism suggests an intentional stance (Dennett 1981); the ascription of "aboutness" to a source of behaviour.

Such an understanding of musical cognition derives from Hendriks-Jansen's notion of *interactive emergence* (Hendriks-Jansen 1996), whereby cognitive experience (including emotion), learning and intelligence are

considered the products of emergent structure in the relationship between interacting behaviours. It also echoes Karmiloff-Smith's account of the development of mind through self-redescription (Karmiloff-Smith 1992), and Bogdan's evolutionary account of the formation of mind through the understanding of other minds (Bogdan 1997).

Modelling Emergence

Crutchfield (1994) presents an understanding of emergence in a computational context which acknowledges the situatedness of the phenomenon. Emergence is seen as a function of a description or explanation of a cognised behaviour, and the constraints of describing system. As a description is elaborated and qualified, it might reach the limits of the system, necessitating more concise or general redescription. This is the understanding adopted in the present model. The limits of the system's descriptive power can be virtual as well as real, providing a powerful tool for characterisation and composition.

The Model

A dynamical systems approach seems most appropriate to a system in which time-evolution and external conditions and environment are critical (Van Gelder & Port, 1995). In a *phase space* consisting of the possible states of the music, with a dimensionality of at least its number of variable parameters plus time, structure of any sort can be represented as an attractor; a trajectory of unique states with some degree of self-similarity. Major changes in the behavioural world of the music correspond to *phase transitions* in the system.

The capacity for self-redescription would imply of a unitary system that it already accounts for its own reformulation, and internal emergence would not strictly be possible. The system at any moment would be an incomplete description of an implicit super-system. Instead, the whole is therefore divided into behaviours and describing agents. This internal division permits something of the dialogue which it was suggested above might be essential to engaged musical activity, and the two elements are here discussed separately.

Behaviours

By analogy with the above axiom, a theory (whether a cognitive theory or a composition) likewise presupposes embodiment in a technology (which may be a body or an instrument), and situatedness in the environment of its activity. Personal musical activity is the result of the interaction of a specific embodiment of a musical object with a particular environment. The interactive performance nature of this project is a discipline for adhering to this principle. At the lowest level, musical "behavors" are situated, embodied agents (i.e. contingent, constrained and aware), interacting hierarchically and acting in accordance with present circumstances and ideas, not prior plans.

To reflect the hierarchical and parallel nature of music, a Complex Adaptive Systems model has been adopted (Holland 1995) - a population of micro-behaviours that can multiply, be active, dormant or die, be transformed, and aggregrate dynamically with other behaviours. Their generation, the level and nature of their activity are governed by the environment they inhabit - a changing landscape formed by the other behaviours and the "outside world" (live musicians in this case). Micro-behaviours generated are evaluated and tagged in the light of the behaviour of the musician in the current performance, and in terms of a set of basic cognitive constraints. Each behaviour has a capacity for self-interaction, described below. In the basic model, their activity takes place within a simple time - frequency - intensity space, taking account of rhythm by sampling frequency at small time intervals.

In its simplest form, the landscape consists of a set of sites in n-dimensional space, representing, for example, values of melodic interval, duration and dynamics. In accordance with thresholds set by the composer, agents develop at these sites, born of and fuelled by the incoming musical material. In as far as they have the resources, these agents compete to be allowed to respond to that input, and in the process form hierarchically-structured melodic, harmonic or rhythmic aggregates in accordance with the strength of their bid and their respective tags - state characteristics reflecting their initial birth and current associations. The pitch input is biased in accordance with Schellenberg's rationalisation of Narmour's implication-realization model (Schellenberg 1994, Narmour 1990), to even the playing field until the actions of composer, performer or the describing system begin to shape it. In its current implementation, the output of this system is sent as MIDI data to Max/FTS to control sound processing and synthesis parameters. The description system can change this landscape by controlling the resourcing and thresholds of certain areas, by seeding or inhibiting agents or aggregates, or by filtering their behaviour.

Description

The mode of description has to be considered in terms of its capacity to generate information which is relevant to the performance function of the system, its predictive power, and descriptive and computational efficiency. The dynamically hierarchical nature of music presents corresponding difficulties for a predictive description of the behaviour of a musical system. Not only must it strive to encapsulate the re-aggregation and scaling discussed above, but it must do so in terms that have both general and specific validity. A single event is a poor indicator of its successor, especially if we consider a uniformly sampled time-series (i.e. wherein most values are not new events). On the other hand, generalised or statistical indicators only relate to general states, and "fuzzy" values stand little chance of being appropriate. In performance terms, music is a succession of *specific* events; the result of a narrative of specific, contextual interactions, possibly widely

distributed over time. The need for specific values becomes yet more acute in the machine context.

Considered as a dynamical system, a behaviour can have only one possible subsequent state; that is, the system would be fully deterministic if all the parameters were known (including those of the performer!). The task of the description in reconstructing a dynamical system is to capture the crucial parameters preceding points of bifurcation - moments of selection between forking paths. The problem is not only to find attractors in the phase space of the system, but to identify an appropriate dimensionality and mapping for the phase space itself. The incompleteness of the system's knowledge of the performer (or composer, over a longer timescale) guarantees the need for redescription, affording interaction and the possibility of evolution.

Linear description techniques hve some validity in a highly correlated musical context such as Western tonality; less so in a language of arbitrary syntax and complexity, particularly in an improvising situation. Not only is the sampled musical surface nonlinear in its evolution, but the significant structure it contains is also nonlinear in the development of its dimensionality. Both the description of a dynamical system by genetic algorithm (Packard 1990) and state space reconstruction (Kantz and Schreiber 1997) have been used, but in this real world, real-time context the necessary arbitrary constraints compromise the evenness of the system's performance.

A connectionist approach to issue of behaviour description was ultimately selected. The simple recurrent networks described in (Elman et al, 1996) embody a form of non-representational memory of the evolution of the system, by feeding the current state back with new input; a phase-space filter, in effect. Such a network is fed both the live performance and the output of the "behaving" system described above. By monitoring the changes in the weights of the network, the relevant dimensionality of a particular activity (analogous to musical structural hierarchy, or the aggregate behaviours within a population of autonomous agents) can be judged, and the "meaning" or mapping of the state of the network for the musical behavioural landscape changed accordingly. The results of cluster analysis of the network weights are used calculate control parameters for the behavioral landscape; filtering, constraining or distorting its relationship with its environment, and controlling the type and rate of its resource generation.

Redescription

The assigning of arbitrary numbers of parallel describing agents in earlier versions of this system raises difficult cognitive issues. Presumably the number of *conscious* agents posited must be limited, and yet arbitrary restrictions on the observation of significant inner structure may discourage the interactive emergence sought. In fact, Crutchfeld's computational emergence implicitly takes account of this by relating the moment of redescription to the descriptive limits of the particular computational context. If this implies a cognitive theory (which he does not claim), its implementation is by emulation, not simulation.

In the current version of this system, redescription takes one of three forms:

i) the re-initialising or major reconfiguration of the behavioural landscape.

ii) the resetting of the weights of the network (re-seeding or setting to vectors stored at critical moments of previous runs).

iii) the running of a parallel network which is constantly learning, and to which the weights of the landscape-managing network are set at points of redescription.

The redescription itself can be prompted by two circumstances:

i) observing sufficiently strong indicators of impending emergence in the observed behaviour. This may be a function of an unsatisfactory description, in which case "realisation" is a more appropriate analogy than emergence.

ii) reaching the limit of the system's capacity (real or virtual) to add to the complexity of its current description. The control of artificial limits is a means of managing the degree of redescription of the whole system, an important factor in its perceived energy level. In the case of a fixed-architecture network, the computational complexity clearly does not increase with the "knowledge" represented. Instead, a difference function is maintained, based on prediction error with respect to certain aspects of the performance. Again, the constitution of this function becomes a compositional decision. A threshold can thus be set - dynamically, if required - which triggers the redescription.

This difference is analogous to the "discrepancy signal" of Linear Prediction Coding, and reflects the "noise" unaccounted for by the previous description. This constitutes a self-environment - a sea of noise (Serres, 1995), or a heat bath - which is the initial source of energy and conditions for the newly-expressed behaviour, as it interacts with its own "differend".. This "self-environment" thus mirrors the relationship between performer and system/composition.

By understanding each identified behaviour as a dynamical system, emergence can be seen as a process of phase transition. Taking a phase portrait view of each strand or aggregate of musical behaviours, the indicators of emergence, of impending redescription, are the precursors of phase transition - the change to a new attractor or range of possibilities for that behaviour (subsystem). These include an increased criticality, or sensitivity to parameter changes, and can thus be "forced" by interaction of the behaviour is sensitive to some aspect of the musician's performance as an order parameter. It is important to note that the range of possibilities for a behaviour is limited dynamically by external parameters (human interaction and the behaviour of other aggregates), and its own state at the moment of redescription. The initial conditions of a

redescribed behaviour determine its activity in the case of a multistable description.

The hierarchy and scaling of complex structures can seen as analogous to those qualities of music identified elsewhere as the keys to a cognitive-based analysis. The self-redefinition of interactive emergence parallels the way in which emotional response itself reshapes the mechanisms of response. The musical object might then be an active extension of working memory, to some degree common to different participants, the response of the whole being reshaped by responses not only present but past. These responses are themselves the result of interactions which may be distributed through space, time and artefacts.

Further Developments

The descriptive functions of the system are currently being re-implemented using a Time-Delay Neural Network (Waibel et al. 1989), in order to improve the accuracy of the reflection of time, as well as performance efficiency. This would also allow the compositional manipulation of the passage of "system experiential" time, possibly relative to some parameter of the musician's performance.

Several compositions have been realised with the system as described here. At present, it requires the "hand-tuning" of parameters, to avoid a lengthy learning process which would soon be overtaken by the evolution of other works or techniques. The designing, constraining and mapping of the relationships described above is a powerful tool in the generation of complex, temporally situated musical surfaces, dependent on local conditions and events. This process of computational sculpting could itself usefully be subject to description in similar terms to that of its performance or instantiation. Indeed, this tuning process could be considered an essential part of the narrative of composition itself. In general, the techniques and materials of a composition develop together in some sort of dialogue. The pre-teaching of certain material to the system would be simple to imagine. However, a more interesting development might be observe the evolution of the weights of the describing network as it moves towards critical moments during the composition phase, in order to be able to induce certain types of phase shift in the piece as it is performed.

The Evolution of Emotional Space

The internal space of the musical object is thus reshaped by the process of interaction. By virtue of this object's role in a reciprocal, feedback relationship with the musician, such a reshaping can be seen as mirroring or at least analogous to some parallel change in the state of the musician's engagement. The present hypothesis is that such a relationship also pertains in a non-realtime, non-computational context. From the standpoint of interactive emergence referred to above, or that of an *enactive* view of cognition (Varela et al, 1991), it is *only* through such relationships that phenomena such as emotion might be studied. Emotion, by this count, is the tip of an iceberg of engagement, however abstract, distant or self-generated the object. Music, for all its apparent incommensurability of labelling, might be a more productive context in which to study and model the *dynamics* of emotion than the compound complexity of interpersonal relationships.

By viewing the entire musician - musical object space as a dynamical system, developing over time, we have a picture of the evolution of a particular instance of musical engagement. This is in effect a continuous cycle of cognition of structure by the ascription of intentionality; of internalisation and externalisation, from the human point of view. Assuming at least the initial conditions of this system to be determined by some sort of emotional disposition - one could argue that this intentionality-ascription constitutes a form of emotional investment - then we can consider the system itself as representing a process of emotional development. Any area of state space which was not initially accessible can be regarded as having been "learned". Without any pretence to the identification of individual emotions, we can see in general terms how attractors in such a system might indicate types of evolution of relationship: obsession, confusion, unstable sensitivity, openness to novel structure.

Port and Van Gelder (1995) suggest that dynamical systems are the most appropriate mode of understanding cognition. If both this dynamical hypothesis and the notion of interactive emergence have validity, learning, or at least the aquisition of experience, might be a form of "tuning" to the indicators of impending phase transition. Such indicators are described differently according to context (Badii and Politti, 1997; Kelso, 1995), but are generally expressed in terms of increasing parametric complexity - the precursor of the need for redescription and emergence. A combination of experience and prediction (simulation, if we're dealing with an internal model) would then allow the evaluation of the current dynamics in terms of its implications for the current state: stasis, termination, degrees of change, revolution, simplification. The emotional significance of these dynamics is thus a function of one's relation to the present state - degrees of pleasure and attachment - and the sequence of emotional investment implicit in the identifying of individual musical behaviours.

References

Badii, R. and Politi A. 1997. *Complexity: Hierarchical structures and scaling in physics.* Cambridge: Cambridge University Press.

Bogdan, R. 1997. *Understanding Minds.* Cambridge MA.: MIT Press.

Bregman, A. 1993. *Audio Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA.: MIT Press.

Crutchfield, J. 1994. The calculi of emergence: computation, dynamics and induction. *Physica D* 75: 11-54.

Dennett, D. 1981. *True Believers: The Intentional Strategy and Why It Works*. Reprinted in J. Haugeland, ed. 199).

Mind Design II: Philosophy, Psychology, Artificial Intelligence. Cambridge, MA.: MIT Press.
Elman, J. L.; Bates, E.A.; Johnson, M.H. ; Karmiloff-Smith, A.; Parisi, D.; and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge MA: MIT Press.
Hendriks-Jansen, H. 1996. *Catching Ourselves in the Act.* Cambridge, MA.: MIT Press.
Holland, J. 1995. *Hidden Order - How Adaptation Builds Complexity.* Reading, MA.: Addison Wesley.
Impett, J. 1994. A Meta-Trumpet(-er). In Proceedings: 1994 International Computer Music Conference. San Francisco: International Computer Music Association.
Kantz, H.and Schreiber, T. 1997. *Nonlinear time series analysis.* Cambridge: Cambridge University Press.
Karmiloff-Smith, A. 1992. *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA.: MIT Press.
Kelso, J. A. S. 1995. *Dynamic Patterns: The Self-Organisation of Brain and Behaviour.* Cambridge, MA.: MIT Press.
Lerdahl, F., and Jackendorf, R. (1983). A Generative Theory of Tonal Music. Cambridge MA: MIT Press.
Meyer, L. B. 1956. *Emotion and Meaning in Music.* Chicago: University of Chicago Press.
Narmour, E. 1990. *The analysis and cognition of basic melodic structures: the implication-realization model.* Chicago: University of Chicago Press.
Packard, N. 1990. A genetic learning algorithm for the analysis of complex data. *Complex Systems* 4, no.5: 543-572.
Schellenberg, E. G. 1994. Expectancy in melody: tests of the implication-realization model. *Cognition* 58: 75-125.
Serres, M. 1995. *Genesis.* Trans. G. James & J. Nielson. Ann Arbor: University of Michigan Press.
Van Gelder, T. and Port, R. F. 1995. It's About Time: An Overview of the Dynamical Approach to Cognition. In R.Port and T. van Gelder (eds.), *Mind as Motion.* Cambridge MA.: Bradford Books/MIT Press.
Varela, F. J.; Thompson,E.; and Rosch, E. 1991. *The Embodied Mind.* Cambridge MA: MIT Press.
Waibel, A.; Hanazawa, T.; Hinton,G.; Shikano, K.; and Lang, K.J. 1989. Phoneme Recognition Using Time-Delay Neural Networks. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol 37, no. 3, 328-339.